Renewable Energy Bidding Strategies Using Multiagent Q-Learning in Double-Sided Auctions

Wei-Yu Chiu, *Member, IEEE*, Chan-Wei Hu, and Kun-Yen Chiu

***Abstract*—Owing to energy liberalization and increasing penetration of renewables, renewable energy trading among suppliers and users has gained much attention and created a new market. This article investigates a double-auction scheme operated by an aggregator with limited supervision for energy trading. To ensure beneficial bidding for renewable generators and end users (EUs) considered as agents, a multiagent Q-learning (MAQL) based bidding strategy is developed to maximize their cumulative reward. Each agent first provides their information about renewable supply or demand to an aggregator who will then return the information about aggregate supply and demand. Without knowing the business model of the aggregator, the agents use Q-tables to estimate the expected cumulative reward and determine their bidding prices accordingly. Finally, the aggregator coordinates energy trading between agents who will then update their Q-tables on the basis of the amount of power bought or sold at the prices they bid. The proposed approach can avoid some unnecessary or unrealistic assumptions generally made by model-based approaches, such as the assumption on the knowledge of others' bidding profiles or the assumption of an oligopoly; it can consider the influence of bidding strategies on the market, which cannot be properly addressed by a conventional proportional allocation mechanism. A numerical analysis using real-world data and considering a profit maximization model for the aggregator shows that the proposed approach outperformed comparable methods in terms of profits of renewable generators and energy satisfaction level of EUs: iterative double auction by approximately 29%, heuristic bidding by 39.9%, random bidding by 38.1%, NSGA-II-based multiobjective approach by 62%, and MOEA/D-based multiobjective approach by 83.1% on average.**

***Index Terms*—Double auction, energy aggregator, multiagent Q-learning (MAQL), multiagent systems, renewable energy trading.**

I. Introduction

ENERGY liberalization and recent technological advancement in electricity markets have created much interest in energy trading through bidding mechanisms [1], [2]. Among various mechanisms, peer-to-peer trading [3] has attracted much attention because it can provide much freedom to market participants. For instance, the participants are allowed to determine their own trading prices, thereby improving price elasticity [4]. Peer-to-peer energy trading seems to be promising, yet challenging at the same time. Because a peer-to-peer platform is a trustless system with no mediating agent, it can be an arduous task to optimize the decision making process for various trading parameters with a huge number of participants [5]. In addition, peer-to-peer trading may reveal participants' information to public or a third party. That information can be further used to deduce the participants' living habit or lifestyle, raising a privacy concern.

Aggregate trading can address a few issues faced by peer-to-peer trading. In an aggregate trading scheme, an aggregator is an impartial mediator to foster energy trading [6]–[9]. An aggregator can, for example, ensure data privacy in end-to-end communication and data integrity during aggregation process; avoid end-to-end communications that may incur system delays when a large number of end users (EUs) and power generators are involved; and reduce price risks for small generators and undertake EUs' task of selecting generators. In general, an energy aggregator groups participants such as renewable generators and EUs in a power system into a single entity and is responsible for various tasks such as information management, service bundling, matching participants, market clearing, and transaction guarantee [10].

However, market participants under the coordination of an aggregator can have limited control over market parameters such as the market prices for buying or selling energy [11]. In certain cases, an aggregator has full control over both the prices and energy quantity to be bought or sold. One example is the use of a contract theory approach for small-scale renewable energy trading that maximizes the revenue of an aggregator [12]. To incentivize power suppliers and consumers to join this aggregate market, individual rationality constraints, incentive compatibility constraints, and opportunity cost constraints are imposed. Consumers then follow the contract items including the price and power quantity determined by the aggregator to meet their power demand.

A double-auction scheme operated by an aggregator with limited supervision seems to be a better alternative. Like a peer-to-peer bidding scheme, power generators and EUs can have freedom and control over their bidding prices and power quantities to maximize their individual utility functions. Like an aggregate energy trading scheme, the aggregator determines the trading quantities and coordinates the trading process, facilitating energy service based business models. In the literature, however, when double auction has been investigated, the primary focus is generally the auction scheme itself instead of bidding strategies of generators and EUs. In [2], for example, authors claimed that any preferred bidding strategies that maximize the

Manuscript received July 16, 2020; revised November 17, 2020; accepted February 3, 2021. Date of publication March 15, 2021; date of current version March 24, 2022. This work was supported by the Ministry of Science and Technology of Taiwan under Grant MOST 109-2221-E-007-020. *(Corresponding author: Wei-Yu Chiu.)*

The authors are with the MOCaRL Lab, Department of Electrical Engineering, National Tsing Hua University, Hsinchu 300044, Taiwan (e-mail: chiuweiyu@gmail.com; huchanwei1204@gmail.com; chiu.kun@gapp.nthu.edu.tw).

Digital Object Identifier 10.1109/JSYST.2021.3059000

benefits of generators can be incorporated into their hierarchical market structure, but only the base and minimum offers of bidding prices were examined. Some studies even employed uniform bidding strategies to evaluate the developed auction scheme [13], [14].

Meanwhile, one bidding agent, a micorgrid central controller in particular, has been considered in an energy market [15], [16]. Extensions from one bidding agent to multiple ones are possible, such as using a proportional allocation mechanism, but agents are generally modeled as price takers in the sense that their bids have little or no influence on the resource price of the market [17]. From this perspective, multiple agents' bidding strategies having an impact on market behaviors in a double-auction scheme need further investigation.

To fill this research gap, this article considers the development of bidding strategies of renewable generators and EUs in a double-auction scheme operated by an aggregator. Both generators and EUs are considered as agents that are treated equally to foster the market activities. To maximize the applicability of the proposed methodology, the business model of the aggregator is considered as unknown to the agents. During the bidding process, EUs and renewable generators can only use aggregate information about demand and supply, i.e., minimal knowledge, to generate bidding prices and quantities without knowing others' bidding profiles.

The scenario considered in this article brings much difficulty to existing model-based approaches that have required some knowledge of the model or made some additional assumptions about the market behaviors in order to formulate bidding strategies for renewable generators and EUs. In [18], for example, a game theoretic approach was applied and a Nash equilibrium was investigated using linear supply functions, but others' bidding profiles must be assumed to be known. In [19], bidding prices were assumed to drop linearly with the increasing level of demand. Although a Stackelberg game approach has gained much attention recently [20], [21], a leader–follower framework is often assumed and the energy market is modeled as an oligopoly, jeopardizing the market parity between generators and EUs [17]. To some extent, model-based approaches can also suffer from their dependency on model uncertainty and can be computationally expensive [22].

Considering the complexity, dynamics, and uncertainty of the electricity auction market, model-free reinforcement learning algorithms have advantages over model-based approaches because of their ability to address nonstationary and stochastic tasks [23], [24]. Using a reinforcement learning algorithm, an agent in the market can learn from experience and maximize their profits in a long run. To realize beneficial bidding to both renewable generators and EUs in consideration of intermittent and stochastic properties of renewable energy sources, this article develops a multiagent Q-learning (MAQL) based bidding strategy, involving dedicated designs of states, actions, and rewards.[1] A renewable generator or an EU is considered as a bidding agent. In the proposed approach, the state of a renewable generator is the difference between the predicted renewable generation and total power demand; the action is the bidding price for selling power to the aggregator; and the reward is the revenue generated by the action. The state of an EU can be either a scalar or a 2-D vector. The difference between local power demand and predicted power generation from all renewable generators is used as a state. If the EU possesses an energy storage system, then the state of charge is the other state. The action is the bidding price for buying power from the aggregator. The reward is a ratio that reflects the satisfaction level for meeting a basic renewable demand and pertains to the energy cost.

Using the proposed MAQL framework, an agent first observes the states and then determines its price and power quantity for bidding. After the agent is notified of the bidding result, i.e., the amount of power that can be sold or bought at the bidding price, it adjusts its pricing policy. This experience is mathematically learned by the update of a Q-table that indicates the value of staying at a particular state-action pair. Given a state, the action that maximizes the Q-value is often preferred. Furthermore, the proposed framework can readily generalize to a more complicated situation in which an infinite number of states are involved, if necessary. This can be done by function approximation: parameterizing the value function using a weight vector. The update of Q-values in a Q-table becomes the update of the weight vector.

The main contributions of this article are as follows. We develop the MAQL-based bidding algorithms for renewable generators and EUs with dedicated designs of states, actions, and rewards in a double-sided auction market coordinated by an aggregator. The algorithms can address the uncertainty of renewable generation and learn to benefit the generators and EUs without knowing the business model of the aggregator. While the impact of bidding agents on market prices has not been modeled in some existing approaches, this article addresses the interactions between agents and their influence on market prices during the bidding process. In contrast with model-based approaches that require some knowledge of the model or make some additional assumptions, the proposed model-free approach can formulate bidding strategies using minimal information; bidding agents adopting the proposed strategies only use information about aggregate renewable demand or supply without knowing others' bidding actions. Finally, a numerical analysis using real-world data is performed to show that the proposed MAQL-based bidding strategies outperform comparable methods in terms of profits of renewable generators, energy costs and satisfaction level of EUs, and price elasticity.

The rest of this article is organized as follows. Section II describes the renewable energy market, relevant models, and associated problem formulation. The proposed MAQL-based bidding strategies are developed in Section III. In Section IV, the proposed methodology is compared with existing methods using real-world data. Finally, Section V concludes this article.

II. System Models and Problem Formulation

This section describes the system operating models of a real-time renewable energy market consisting of renewable generators, EUs, and an aggregator. In each time slot t, both

[1] R.O.C Patent No. I687890 (2020)

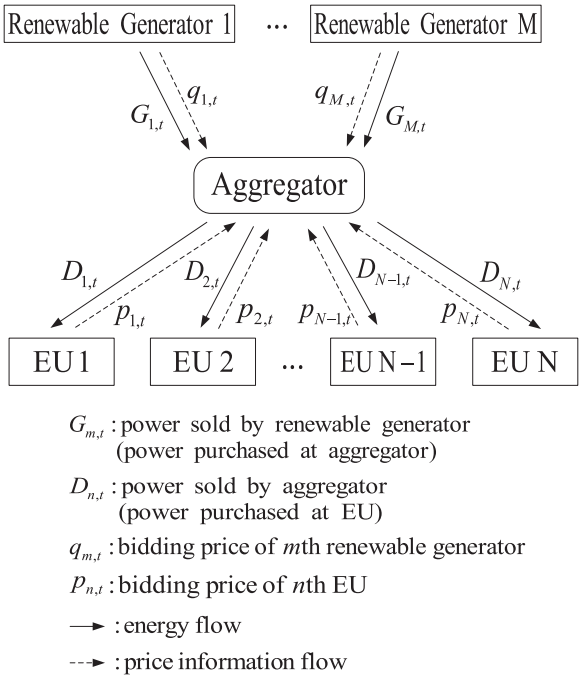


Fig. 1. Structure of the real-time renewable market. The aggregator can purchase power from renewable generators and sell it to EUs in response to bidding prices.

TABLE I
NOTATION

Symbols	Descriptions
t	Time index
M	Total number of renewable generators
N	Total number of EUs
m	Index of a renewable generator
n	Index of an EU
$q_{m,t}$	Bidding price of mth renewable generator at t
$q_m^{\max}$	Maximum bidding price of mth renewable generator
$q_m^{\min}$	Minimum price of mth renewable generator
$p_{n,t}$	Bidding price of nth EU at time t
$p_n^{\max}$	Maximum affordable price at EU n
$p_n^{\min}$	Minimum price for making a bid at EU n
$G_{m,t}$	Power sold from mth renewable generator at time t
$\hat{G}_{m,t}$	Predicted power generation of mth renewable generator at t
$G_{m,t}^{\max}$	Power generation of mth renewable generator at t
$D_{n,t}$	Power purchased at nth EU at t
$D_{n,t}^{\text{base}}$	Basic renewable demand at t
$\hat{D}_{n,t}$	Desired renewable demand of nth EU at t
L_m^{res}	Maintenance cost of mth renewable generator
$\beta_{n,t}$	Battery level of nth EU at t
$\beta_n^{\min}$	Minimum energy level of the battery
$\beta_n^{\max}$	Maximum battery capacity
$r_{n,t}^{\text{EU}}$	Reward function of nth EU at t
$r_{m,t}^{\text{gen}}$	Reward function of mth generator at t

renewable generators and EUs submit their bids to the aggregator. The renewable generators desire to sell power at high prices, and EUs aim to buy a suitable amount of power at low prices. The aggregator dictates the amount of power to be exchanged in order to maximize its own profit. A real-time scheduling problem is considered. Fig. 1 depicts the renewable energy trading scheme in which several renewable generators and EUs are connected to an aggregator [11]. Related models and our problem formulation are described in the following sections. Table I summarizes the notation used throughout this article.

A. Renewable Generators

In a real-time electricity market, power demand depends on EUs' behaviors, and power supply from renewable generators varies with changing environments, such as the weather conditions and sunshine duration, incurring generation uncertainty. Let $G_{m,t}^{\max}$ denote the actual power generation and $\hat{G}_{m,t}$ denote the predicted power generation. The generation cost for the mth renewable generator mainly comes from the maintenance and installation cost, which can be considered as independent of supply quantities and treated as a constant L_m^{res}. The mth renewable generator determines price $q_{m,t}$ for selling the predicted power generation $\hat{G}_{m,t}$, and the aggregator determines the amount of electricity $G_{m,t}$ bought from the generator. Thus, the power $G_{m,t}$ sold by the generator is a function of $q_{m,t}$ and is upper bounded by the predicted generation $\hat{G}_{m,t}$, i.e.,

$$0 \le G_{m,t}(q_{m,t}) \le \hat{G}_{m,t}. \tag{1}$$

From the perspective of a renewable generator, an analytical form of $G_{m,t}$ is not available; the generator only knows the value of $G_{m,t}$ after submitting the price $q_{m,t}$ in each time slot t. To simplify our notation, we will omit the controlling variable $q_{m,t}$ from the controlled variable $G_{m,t}$.

The optimal bidding price for the mth renewable generator can be achieved by maximizing its profit as follows [7]:

$$\max_{q_{m,t}} \quad \mathbb{E}\left\{\sum_{t=1}^{T} q_{m,t}G_{m,t} - L_m^{\text{res}} - U(q_{m,t})\right\} \tag{2}$$
$$\text{subject to} \quad q_m^{\min} \le q_{m,t} \le q_m^{\max}$$

where $\mathbb{E}\{\cdot\}$ represents the expectation with respect to the random variable $G_{m,t}$, and $q_m^{\min}$ and $q_m^{\max}$ represent the minimum and maximum bidding prices, respectively. The penalty function $U(q_{m,t})$ for generators in (2) can be calculated as

$$U(q_{m,t}) = \{q_m^{\max}(G_{m,t} - G_{m,t}^{\max}), 0\} \tag{3}$$

which is the penalty paid to the aggregator if the generator cannot provide power it promises to deliver to the aggregator. If the penalty occurs, we have

$$G_{m,t} > G_{m,t}^{\max}. \tag{4}$$

To some extent, the penalty can be interpreted as the imbalance settlement pricing in a spot market [25].

If $G_{m,t}$ was deterministic and its explicit form is given to renewable generators, then (2) would be an optimization problem. In our scenario, however, $G_{m,t}$ is only available after price $q_{m,t}$ is submitted in each time slot t, so (2) is a learning problem. The minimum price $q_m^{\min}$ can be related to feed-in tariff prices; in such a case, there is no reason for selling power at a price that is less than a feed-in tariff price. If $q_{m,t} < q_m^{\min}$, the renewable generator can simply sell all its generated power to the market

regulated by a local government with the feed-in tariff price, which is more beneficial than staying in the trading market coordinated by an aggregator.

B. End Users

Suppose that N EUs are involved in the real-time bidding market, some of which are equipped with an energy storage system, a battery in particular. Let $\beta_{n,t}$ denote the energy level at time t. The battery dynamics can be expressed as

$$\beta_{n,t+1} = \begin{cases} \beta_{n,t} + \eta_c \Delta\beta_{n,t}, & \Delta\beta_{n,t} \geqslant 0 \\ \beta_{n,t} + \eta_d \Delta\beta_{n,t}, & \Delta\beta_{n,t} < 0 \end{cases} \quad \text{subject to} \quad \beta_n^{\min} \leq \beta_{n,t} \leq \beta_n^{\max} \tag{5}$$

where $\Delta\beta_{n,t}$ is positive/negative for a charging/discharging event, $\beta_n^{\min}$ and $\beta_n^{\max}$ represent the minimum and maximum capacity, and $\eta_c < 1$ and $\eta_d > 1$ represent the charging and discharging efficiency, respectively.

Let $D_{n,t}^{\text{base}}$ denote the basic renewable demand at time t and $\hat{D}_{n,t}$ denote the desired demand, where $\hat{D}_{n,t} \geq D_{n,t}^{\text{base}}$. The basic renewable demand $D_{n,t}^{\text{base}}$ is related to the required quota of green consumption, and the desired demand $\hat{D}_{n,t}$ is related to $D_{n,t}^{\text{base}}, \beta_n^{\min}, \beta_n^{\max}$, and $\beta_{n,t}$.

The nth EU determines price $p_{n,t}$ for purchasing the desired renewable demand $\hat{D}_{n,t}$, and the aggregator determines the amount of power $D_{n,t}$ sold to the EU. Thus, the power $D_{n,t}$ purchased from the aggregator is a function of the bidding price $p_{n,t}$ offered by the nth EU and is upper bounded by $\hat{D}_{n,t}$, i.e.,

$$0 \leq D_{n,t}(p_{n,t}) \leq \hat{D}_{n,t}. \tag{6}$$

To simplify our notation, we will omit the controlling variable $p_{n,t}$ from the controlled variable $D_{n,t}$. In general, increasing the price $p_{n,t}$ increases the chance of making $D_{n,t}$ close to $\hat{D}_{n,t}$.

The goal of each EU is to minimize the total expenditure while the basic renewable demand is met in consideration of the battery status if any. The objective function of an EU can be mathematically expressed as

$$\max_{p_{n,t}, \hat{D}_{n,t}, \Delta\beta_{n,t}} \quad \mathbb{E}\left\{ \sum_{t=1}^{T} \frac{D_{n,t}}{\hat{D}_{n,t}} \left(1 - \frac{p_{n,t}}{p^{\max}}\right) \right\} \quad \text{subject to} \quad p_n^{\min} \leq p_{n,t} \leq p_n^{\max} \tag{7}$$

where $\mathbb{E}\{\cdot\}$ represents the expectation with respect to the random variable $D_{n,t}$, and $p_n^{\min}$ and $p_n^{\max}$ represent the minimum price for making a bid and maximum affordable bidding price, respectively. The term $\frac{D_{n,t}}{\hat{D}_{n,t}} \in [0, 1]$ indicates the energy satisfaction level and the term $(1 - \frac{p_{n,t}}{p^{\max}}) \in [0, 1)$ is related to the reciprocal of the energy cost, which is larger when the bidding price $p_{n,t}$ is closer to the minimum price $p_n^{\min}$. Maximizing the objective function in (7) implies more power purchased at a lower price. In our scenario, an EU acts as a flexible load.

C. Aggregator

An aggregator ensures data privacy in end-to-end communication and data integrity during aggregation process, and avoids end-to-end communications that may incur system delays when many EUs and power generators are involved. In our scheme, requests of renewable power demand $\hat{D}_{n,t}$ and supply $\hat{G}_{m,t}$ along with bidding prices $p_{n,t}$ and $q_{m,t}$ are sent to the aggregator. The aggregator then determines the amount of power bought from renewable generators $(G_{m,t}, m = 1, 2, \ldots, M)$ and the amount of power sold to EUs $(D_{n,t}, n = 1, 2, \ldots, N)$ at each time step t in order to balance the supply and demand, i.e.,

$$\sum_{n=1}^{N} D_{n,t} = \sum_{m=1}^{M} G_{m,t}. \tag{8}$$

Several business models have been developed to maximize the profit of the aggregator. For example, strategic curtailment of generation in an electricity market was proposed in [26] by maximizing curtailment profit subject to locational marginal prices and curtailment constraints. A double-auction scheme has been widely used to maximize the social welfare [2]. When the role of market institution in the double-auction scheme is replaced by an aggregator, the social welfare can be considered as the profit for the aggregator. Many variations on the double-auction scheme are also possible. In [27] and [26], nonlinear objective functions in a microgrid controller optimization problem were examined. In [7], the difference between the reward received from electric utility operator and the compensation provided to EUs was maximized.

This article assumes that the business model of the aggregator is unknown to renewable generators and EUs. Under this assumption, the development of bidding strategies for the generators and EUs without considering the aggregator model is the primary focus. Meanwhile, the aggregator can benefit from the proposed market mechanism by maximizing its own profit during the coordination of energy trading.

Because energy trading is coordinated by the aggregator given predicted power generation $\hat{G}_{m,t}$ from renewable generators, it is possible that true power generation is less than the predicted power generation, i.e., $G_{m,t}^{\max} < \hat{G}_{m,t}$. If this occurs, a renewable generator can be asked to provide more power than what it is able to produce, leading to the supply shortage. In this case, the penalty $U(q_{m,t})$ in (3) is given to the generator and the aggregator must compensate this individual supply shortage through various ways. For example, the aggregator may have to use its own energy storage system for the compensation.

III. Proposed Bidding Strategies

This section describes the bidding scheme in which renewable generators and EUs can have their impacts on their own profits and costs while the aggregator can coordinate the energy trading process. The concept is to allow renewable generators and EUs to submit their bidding prices in response to their supply and demand, respectively. Beneficial bidding prices are learned using the proposed MAQL. The bidding procedure is first examined, followed by various designs on states, actions, and rewards of the MAQL in the framework of reinforcement learning.

Fig. 2 presents the bidding procedure consisting of four steps. In the first step, information about individual power supply $\hat{G}_{m,t}$

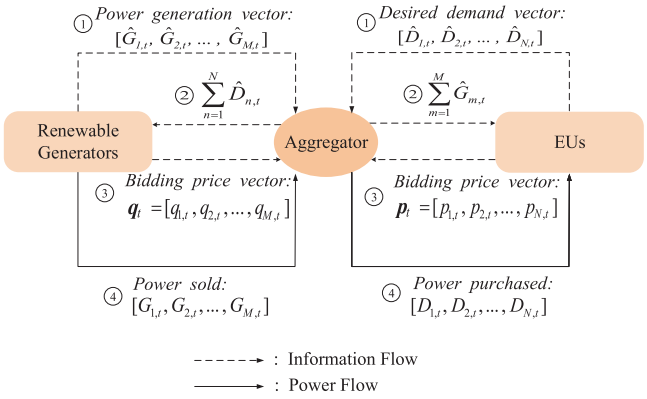


Fig. 2. Information and power flows of the proposed bidding scheme for aggregate renewable energy trading at time t.

and demand $\hat{D}_{n,t}$ is sent to the aggregator. Let us denote

$$\hat{D}_t = \sum_{n=1}^{N} \hat{D}_{n,t} \quad \hat{G}_t = \sum_{m=1}^{M} \hat{G}_{m,t}. \tag{9}$$

In the second step, the aggregator provides information about total power demand $\hat{D}_t$ and total power supply $\hat{G}_t$ to renewable generators and EUs, respectively. Information exchange in the first two steps is coordinated by the aggregator. The aggregator only reveals the aggregate information of demand and supply during the trading process. The individual privacy of power users and suppliers can be preserved to some extent. In the third step, renewable generators and EUs submit desired prices $q_{m,t}$ and $p_{n,t}$ to the aggregator on the basis of the total demand and supply, respectively. This step is similar to peer-to-peer energy trading through a typical bidding process. In the final step, the aggregator buys power $G_{m,t}$ from renewable generators and sell power $D_{n,t}$ to EUs in response to the bidding prices. The following sections present detailed designs for steps 1 and 3 in Fig. 2.

The double-auction scheme and the developed bidding strategies form a distributed control architecture. The aggregator distributes information about aggregate renewable demand and supply to EUs and generators while EUs and generators make their own control decision (bidding prices) individually. In this article, we assume that some formalized contractual agreements exist for the aggregator to match participants and maintain its commitments and obligations. For instance, Universal Smart Energy Framework allows the aggregator to match the contracts between supplying and demanding participants [28]. The information exchange mechanism in Fig. 2 can then be part of the contracts between the participants.

A. Desired Demand in Step 1

In step 1, EUs need to submit their power demand to the aggregator. In our scheme, the EUs are encouraged and designed to purchase the power quantity

$$\hat{D}_{n,t} = D_{n,t}^{\text{base}} + (\beta_n^{\max} - \beta_{n,t}) \tag{10}$$

so as to meet their basic renewable demand and have a fully charged battery, if any. For EUs with batteries, the charging/discharging control $\Delta\beta_{n,t}$ is designed as

$$\Delta\beta_{n,t} = \max\{D_{n,t} - D_{n,t}^{\text{base}}, -(\beta_{n,t} - \beta_n^{\min})\}$$
$$\text{subject to} \quad \Delta\beta_{n,t}^{\min} \le \Delta\beta_{n,t} \le \Delta\beta_{n,t}^{\max} \tag{11}$$

where maximum power $\Delta\beta_{n,t}^{\max}$ and minimum power $\Delta\beta_{n,t}^{\min}$ in charging and discharging activities, respectively, are calculated as

$$\Delta\beta_{n,t}^{\max} = D_{n,t} - D_{n,t}^{\text{base}} \tag{12}$$

$$\Delta\beta_{n,t}^{\min} = \beta_n^{\min} - \beta_{n,t}. \tag{13}$$

In this case, if the power bought from the aggregator is larger than the power needed for the basic renewable demand, then the excessive energy is stored in the battery; otherwise, the insufficient amount of power is compensated by discharging the battery. Theorem 1 shows that the battery constraints in (5) with ideal charing and discharging can be satisfied using (10) and (11).

Theorem 1: For the demand in (10) and battery control in (11) with $\eta_c \approx \eta_d \approx 1$, the battery constraints $\beta_n^{\min} \le \beta_{n,t} \le \beta_n^{\max}$ in (5) are satisfied for all t.

Proof: For discharging, we have $\Delta\beta_{n,t} \ge -(\beta_{n,t} - \beta_n^{\min})$ according to (11). According to (5), we have

$$\beta_{n,t+1} = \beta_{n,t} + \eta_d \Delta\beta_{n,t} \ge \beta_{n,t} - \eta_d(\beta_{n,t} - \beta_n^{\min}) \approx \beta_n^{\min}. \tag{14}$$

For charging, we have

$$\Delta\beta_{n,t} = D_{n,t} - D_{n,t}^{\text{base}}. \tag{15}$$

By (5), (6), and (10), we have

$$\beta_{n,t+1} = \beta_{n,t} + \eta_c(D_{n,t} - D_{n,t}^{\text{base}}) \le \beta_{n,t} + \eta_c(\hat{D}_{n,t} - D_{n,t}^{\text{base}})$$
$$= \beta_{n,t} + \eta_c(D_{n,t}^{\text{base}} + \beta_n^{\max} - \beta_{n,t} - D_{n,t}^{\text{base}}) \approx \beta_n^{\max} \tag{16}$$

which completes the proof. ■

For those EUs without batteries, the associated operation is relatively simple. The desired demand is then set to equal to the basic renewable demand

$$\hat{D}_{n,t} = D_{n,t}^{\text{base}}. \tag{17}$$

B. Learning Algorithms in Step 3

In step 3, a way of submitting bidding prices that benefit individual renewable generators and EUs is required. We propose to use the concept of Q-learning, a reinforcement learning algorithm. Reinforcement learning [29] is a mathematical framework that explores or learns optimal policies to address stochastic and nonstationary tasks. The decision maker, also known as the agent, interacts with an environment by observing the current state and taking an action in a sequence of time steps. At each time step, the agent receives a reward signal from the environment with regard to the execution of the action selected. The agent learns to make decisions in order to maximize the cumulative reward obtained from the environment.

The concept of reinforcement learning can be formalized using a Markov decision process. Let $\mathcal{S}$ be the state space in the environment and s_t denote the state at time t (current state), where $s_t \in \mathcal{S}$. Given s_t, the agent selects an action a_t from the action set $\mathcal{A}$. The transition from s_t to a next state s_{t+1} occurs and the environment returns a numerical reward r_{t+1} used to evaluate the action a_t executed by the agent at state s_t. A policy $\pi(a_t|s_t)$ is a probability function for selecting action a_t given state s_t. The goal of the agent is to find an optimal policy that maximizes the expected cumulative reward from time t. Define an action-value function $Q_\pi(s_t, a_t)$ under policy π as

$$Q_\pi(s_t, a_t) = \mathbf{E}_\pi \left[\sum_{k=t+1}^{T} \gamma^{k-t-1} r_k | s_t, a_t \right] \quad \forall s_t \in \mathcal{S}, \forall a_t \in \mathcal{A} \tag{18}$$

where $\gamma \in (0, 1]$ represents a discount rate. The optimal action-value function $Q_*(s_t, a_t)$ satisfies

$$Q_*(s_t, a_t) = \max_\pi Q_\pi(s_t, a_t). \tag{19}$$

The optimization process of (19) can be achieved using a Q-learning algorithm without knowing state transition probabilities [30], [31].

In Q-learning, an agent uses a Q-table $Q(s_t, a_t)$ to estimate $Q_*(s_t, a_t)$ and has two strategies to select an action: For exploitation, the agent selects the best action at the current state that yields the highest Q-value with probability $1 - \epsilon$; for exploration, the agent gathers more information by randomly selecting actions with probability ϵ. This method is called ϵ-greedy action selection. The update rule for the Q-table is given by

$$\begin{aligned} Q(s_t, a_t) = {} & (1 - \lambda) Q(s_t, a_t) \\ & + \lambda \left(r_{t+1} + \gamma \max_{a_{t+1} \in \mathcal{A}} Q(s_{t+1}, a_{t+1}) \right) \end{aligned} \tag{20}$$

where $\lambda \in (0, 1)$ represents a learning rate. To implement Q-learning as bidding strategies, we need to define our states s_t, actions a_t, and the corresponding rewards r_t.

In our scheme, renewable generators and EUs are considered as agents. For a renewable generator, the state $s^{\text{gen}}_{m,t}$ is designed as the difference between the predicted supply and total demand

$$s^{\text{gen}}_{m,t} = \hat{G}_{m,t} - \hat{D}_t \tag{21}$$

where $\hat{D}_t$ is defined in (9). The bidding price will depend on $s^{\text{gen}}_{m,t}$ that is generally discretized over a range of interest. The state in (21) contains information about multiple agents in an aggregate way, i.e., $\hat{D}_t$; knowledge of individual $\hat{D}_{n,t}$ is not known to the mth generator.

For an EU, a 2-D state $(s^{\text{EU}}_{n,t}, s^{\text{SoC}}_{n,t})$ is designed, one state $s^{\text{EU}}_{n,t}$ for the difference between its power demand $\hat{D}_{n,t}$ and total power supply and the other state $s^{\text{SoC}}_{n,t}$ for battery information, if any. The state $s^{\text{EU}}_{n,t}$ is defined as

$$s^{\text{EU}}_{n,t} = \hat{D}_{n,t} - \hat{G}_t \tag{22}$$

where $\hat{G}_t$ is defined in (9). State $s^{\text{EU}}_{n,t}$ contains information about other agents in an aggregate way as well, i.e., $\hat{G}_t$; knowledge of individual $\hat{G}_{m,t}$ is not known to the nth EU. The other state $s^{\text{SoC}}_{n,t}$ is related to the state of charge (SoC) and defined as

$$s^{\text{SoC}}_{n,t} = \frac{\beta_{n,t} - \beta^{\min}_n}{\beta^{\max}_n - \beta^{\min}_n}. \tag{23}$$

Similarly, $s^{\text{EU}}_{n,t}$ and $s^{\text{SoC}}_{n,t}$ are discretized over a range of interest.

Submitted prices are the actions taken by the agents. Let

$$a^{\text{gen}}_{m,t} = q_{m,t} \in \mathbb{R}^+ \text{ and } a^{\text{EU}}_{n,t} = p_{n,t} \in \mathbb{R}^+ \tag{24}$$

be the actions taken by renewable generators and EUs, respectively. Their action spaces (price ranges) of $a^{\text{gen}}_{m,t}$ and $a^{\text{EU}}_{n,t}$ are discretized over $[q^{\min}_m, q^{\max}_m]$ and $[p^{\min}_n, p^{\max}_n]$, respectively.

From the perspective of renewable generators, they seek to sell power $G_{m,t}$ at high prices $q_{m,t}$ in order to maximize their profits. For generator m at time slot t, the following reward function can be used:

$$r^{\text{gen}}_{m,t} = q_{m,t} G_{m,t} - U(q_{m,t}), \quad \forall t \in [1, T] \tag{25}$$

where $U(q_{m,t})$ represents the penalty if the generator cannot provide enough power for the aggregator as it originally promised. From the perspective of EUs, satisfying the desired power demand and charging the battery, if any, at low prices are the goal. The reward function of the nth EU at time slot t is then designed as

$$r^{\text{EU}}_{n,t} = \frac{D_{n,t}}{\hat{D}_{n,t}} \left(1 - \frac{p_{n,t}}{p^{\max}} \right), \quad \forall t \in [1, T]. \tag{26}$$

In this case, a larger reward generally implies more power purchased at a lower price.

We summarize the steps in the proposed MAQL bidding algorithm as follows. At time step t, the generator and EU receive the reward r_{t+1} from the bidding environment by disclosing their price a_t using ϵ-greedy action selection in current state s_t, and are then transferred to next state s_{t+1}. Next, at time step $t + 1$, the algorithm updates the Q-value for the pair (s_t, a_t) by (20). With the designs of states in (21)–(23), actions in (24), and rewards in (25) and (26), the learning algorithm for price bidding is presented in Algorithm 1. The algorithm is valid for both renewable generators and EUs. For renewable generators, $s_t = s^{\text{gen}}_{m,t}$ in (21), $a_t = a^{\text{gen}}_{m,t}$ in (24), and $r_t = r^{\text{gen}}_{m,t}$ in (25) are used. For EUs, $s_t = (s^{\text{EU}}_{n,t}, s^{\text{SoC}}_{n,t})$ in (22) and (23), $a_t = a^{\text{EU}}_{n,t}$ in (24), and $r_t = r^{\text{EU}}_{n,t}$ in (26) are used. The MAQL algorithm adopts the Q-learning framework and uses states involving information about multiple agents. Because this information is gathered and provided by the aggregator in an aggregate way, privacy of individual renewable generators and EUs can be preserved to some extent.

Remark 1: In Fig. 2, the desired demand in step 1 as expressed in (10) was calculated by each EU; in step 2, information about the aggregate demand and supply in (9) was collected by the aggregator and then sent to individual EUs and generators; in step 3, bidding prices were produced by EUs and generators by running Algorithm 1, including the calculations of (22)–(24), and (26) at EUs and the calculations of (21), (24), (25) at generators; finally, in step 4, the amount of power flow was determined

Algorithm 1: Proposed MAQL for Renewable Generators and EUs.

Input: Learning rate λ.
Output: Learned bidding strategy derived from Q.
1: Initialize $Q(s,a)$ randomly for all $s \in \mathcal{S}$, $a \in \mathcal{A}$. Set $t := 0$.
2: **While** s_t is not a terminal state **do**
3: Choose a desired price a_t using policy derived from $Q(s_t, \cdot)$, i.e., ϵ-greedy action selection.
4: Submit the price to the aggregator and receive a reward r_{t+1}.
5: Update $Q(s_t, a_t)$ using (20).
6: $t := t + 1$
7: **end while**

by the aggregator, for example, through the maximization of its own profit.

Remark 2: Renewable generators and EUs determine their bidding prices using ϵ-greedy action selection, as shown in line 3 of Algorithm 1. They record their outcomes of past bidding in Q-tables, and then select the best prices that maximize the Q-values most of the time accordingly. The best prices depend on the aggregate information about power supply and demand. From time to time (with a small probability of ϵ), they randomly select prices to explore better actions if any.

IV. Numerical Results

This section presents the numerical analysis on the proposed MAQL-based bidding strategies for renewable energy trading. A real-time bidding market consisting of an aggregator, renewable generators, and EUs was examined [32], [33]. For the purpose of numerical analysis, the aggregator maximized its own profit by solving the following [34]:

$$\begin{aligned} \max_{D_{n,t}, G_{m,t}} \quad & \sum_{n=1}^{N} p_{n,t} D_{n,t} - \sum_{m=1}^{M} q_{m,t} G_{m,t} \\ \text{subject to} \quad & 0 \le G_{m,t} \le \hat{G}_{m,t} \\ & 0 \le D_{n,t} \le \hat{D}_{n,t} \\ & \sum_{n=1}^{N} D_{n,t} = \sum_{m=1}^{M} G_{m,t}. \end{aligned} \tag{27}$$

However, renewable generators and EUs were not aware of such a profit maximization model. EUs were randomly selected to be equipped with batteries. Real-world data from European Network of Transmission System Operators for Electricity (ENTSO-E) [35] were used, including renewable demand and renewable generation from biomass (18.43%), hydropower (7.62%), wind offshore (10.86%), wind onshore (44.76%), and photovoltaics (18.33%), to produce $D_{n,t}^{\text{base}}$ and $G_{m,t}^{\max}$. A quarter-hourly resolution and optimization horizon $T = 96$ were considered. Fig. 3 shows the data distribution of average quarter-hourly power generation and demand from 2019 to 2020.

To model the generation uncertainty, we considered a state-of-the-art prediction method of renewable generation from [36]

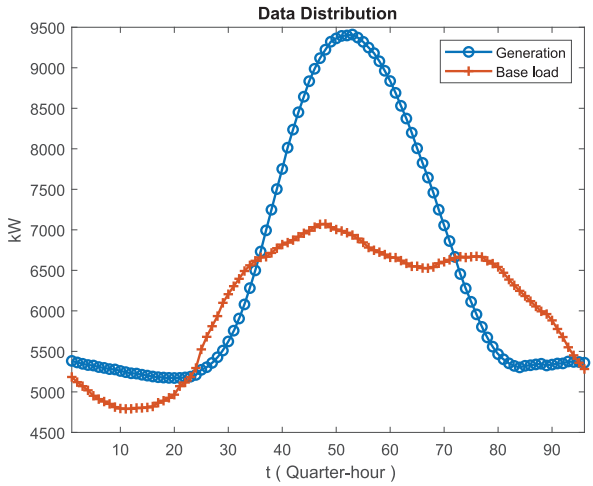


Fig. 3. Distribution of average power generation and demand.

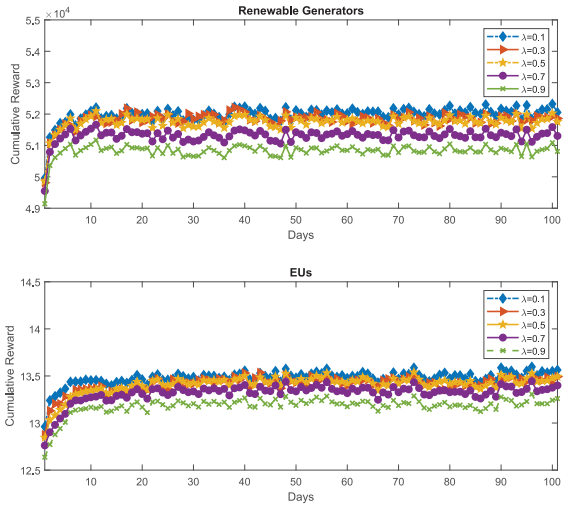


Fig. 4. Cumulative reward of the proposed MAQL algorithm with variations of learning rate from 0.1 to 0.9.

that yielded at most a 5% prediction error on average. As such, $\hat{G}_{m,t}$ was randomly sampled from $[0.95G_{m,t}^{\max}, 1.05G_{m,t}^{\max}]$ in our simulations.

A. Learning Rate Tuning

The learning rate $\lambda = 0.1$ was set in our proposed MAQL algorithm. The learning rate defines how much the updated Q-value learns from the new state-action pair. In general, there is no explicit rule how this parameter should be chosen. Fig. 4 shows the variations of learning rate from 0.1 to 0.9. The proposed learning algorithm performed well when λ was below 0.5.

B. State Design

To illustrate the use of information about total power demand and supply, we examined the following two conditions:

1) generators have or have no information about total demand (EUs have information about total supply);
2) EUs have or have no information about total supply (generators have information about total demand).

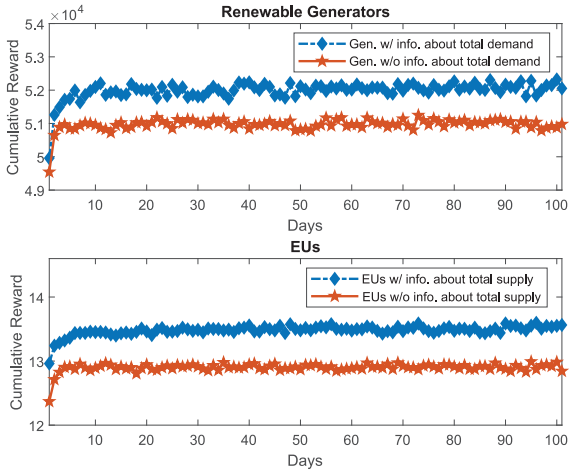


Fig. 5. Cumulative reward of different state designs for the proposed MAQL-based bidding algorithm over 100 days.

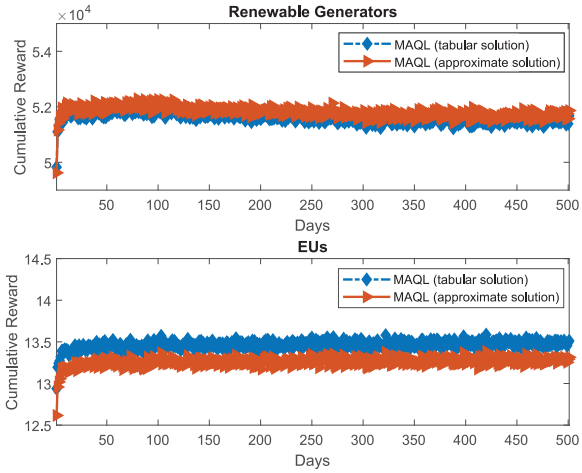


Fig. 6. Cumulative reward of the proposed bidding algorithm using MAQL over 500 days.

Fig. 5 shows that information regarding total power demand and supply was important for participants to formulate a decent bidding strategy. For a renewable generator, the proposed state $s_{m,t}^{\text{gen}}$ was designed as in (21), indicating the difference between the predicted supply and total demand; whereas the comparable state of generator ignoring information about the total demand was $s_{m,t}^{\text{gen-ignor}} = \hat{G}_{m,t}$. For an EU, the proposed state $s_{n,t}^{\text{EU}}$ was designed as in (22), indicating the difference between its power demand $\hat{D}_{n,t}$ and total power supply; whereas the comparable state of EU ignoring information about the total supply was $s_{n,t}^{\text{EU-ignor}} = \hat{D}_{n,t}$.

C. Computation Study of MAQL

It is worth mentioning that the framework presented in Algorithm 1 can readily generalize to approximate solution methods. In that case, the Q function is parameterized by a weight vector that is updated in each time slot using stochastic gradient descent methods. Furthermore, the proposed learning framework can be combined with any existing approaches that are applicable in a real-time market to reduce the time cost if necessary; in line 3 of Algorithm 1, action a_t will be produced by any approach employed. This is possible because the multiagent framework adopts an off-policy method, providing flexibility to use other behavior policies that generate data.

Fig. 6 shows the performance of the proposed MAQL with and without function approximation via tile coding. Both algorithms were trained over 408 days and their learning curves converged approximately in five days. In general, function approximation generalizes better than Q-learning and can learn faster. Nevertheless, our simulation results did not show such an advantage of function approximation. This can be due to the fact that the state space is not large and our state design provides compact information sufficient for an agent to learn well.

Fig. 7 illustrates battery control using (11) in the proposed MAQL-based bidding scheme. When bidding prices were low, batteries were charged for later use. When bidding prices were high, batteries were discharged to reduce the cost. This justified the effectiveness of our battery control design.

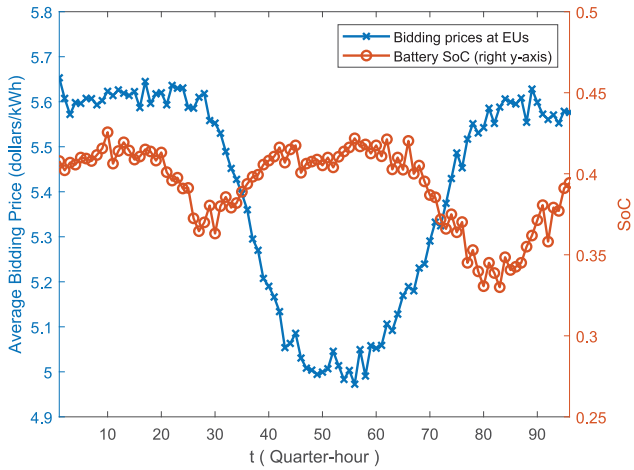
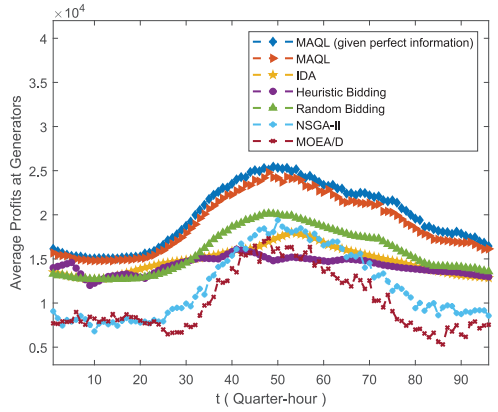


Fig. 7. Average bidding prices at EUs and battery SoC using MAQL. Batteries were charged at low prices and discharged at high prices to reduce the energy cost.

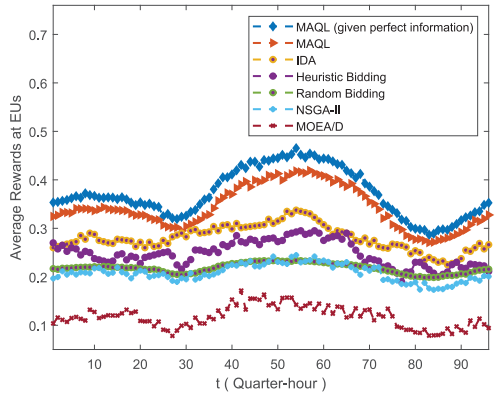
D. Comparison With Existing Approaches

Our approach was compared with the following:

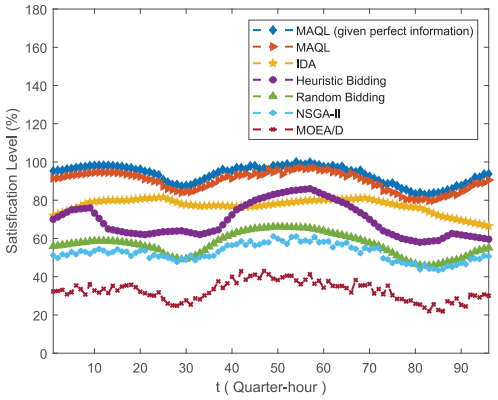
1) an iterative double auction (IDA) that maximized the overall benefit of market participants [27];
2) a heuristic pricing strategy modified from [37] that used a bidding policy represented by a linear function;
3) a random policy submitting a price uniformly at random without using any information;
4) a multiobjective optimization approach that employed evolutionary algorithms to jointly address conflicting objectives of the renewable generators, EUs, and aggregator [33].



(a)



(b)



(c)

Fig. 8. Performance comparison at the renewable generators and EUs given the profit maximization model of the aggregator. The best but ideal performance was attained by the MAQL given perfect information on renewable generation. The proposed MAQL-based bidding scheme in the presence of prediction errors for renewable generation yielded the highest profits for the renewable generators and highest rewards for the EUs, as shown in (a) and (b), and achieved the highest satisfaction level of EUs, as shown in (c).

For the heuristic pricing strategy, renewable generators and EUs tended to bid low prices as power generation and demand increased, which was modeled as monotonically decreasing linear functions. For the multiobjective approaches,

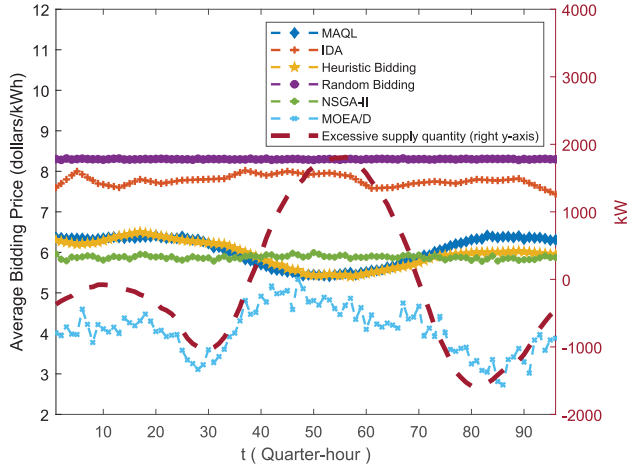
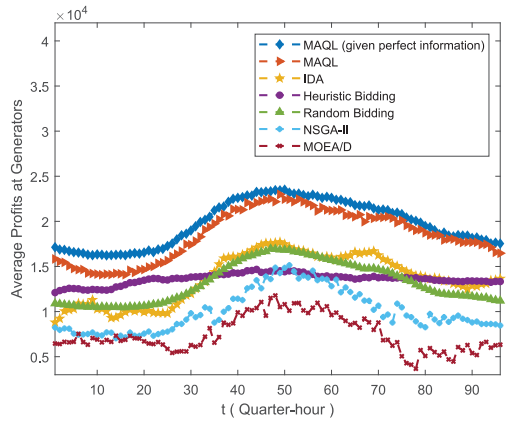


Fig. 9. Bidding prices at EUs with respect to the excessive supply quantity (right y-axis). The excessive supply quantity is positive when there is more supply than demand. Our MAQL-based bidding strategy presented an elastic property; price variations of comparable approaches with respect to the excessive supply quantity presented price inelasticity.

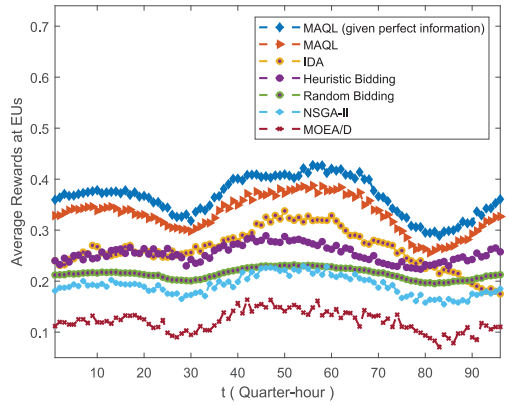
optimization was performed at the aggregator, and decision variables $q_{m,t}$, $p_{n,t}$, $G_{m,t}$, and $D_{n,t}$ were determined by solving the multiobjective optimization problem comprising (2), (7), and the profit maximization problem of the aggregator in (27); batteries, if any, were controlled by the same rule described in (11). Two state-of-the-art multiobjective evolutionary algorithms called NSGA-II and MOEA/D [38], [39] were employed to find Pareto optimal solutions, and the resulting knee solution was selected. For algorithm MOEA/D, prior information was employed to normalize each objective.

Fig. 8 shows the average profits and rewards of the generators and EUs using various methods. The best but ideal performance was attained by our MAQL algorithm given perfect information on renewable generation. In Fig. 8(c), the satisfaction level percentage is defined as $\min\{D_{n,t}, D_{n,t}^{\text{base}}\}/D_{n,t}^{\text{base}}$. The MAQL outperformed IDA, heuristic strategy, random bidding, and multiobjective approaches in the presence of prediction errors for renewable generation. The IDA focused on social welfare optimization, and thus the bidding strategy of EUs was suboptimal. The multiobjective approaches involved a large number of decision variables, including the bidding prices, to be determined at the aggregator, and thus algorithm convergence was difficult to attain. In addition, multiobjective optimization was performed in each time slot rather than over the optimization horizon T, yielding a suboptimal design. By contrast, our learning-based bidding focused on the bidding strategy and addressed fewer parameters than the multiobjective approaches; it learned the data trends and maximized the cumulative reward of the agents, thereby producing a superior level of performance.

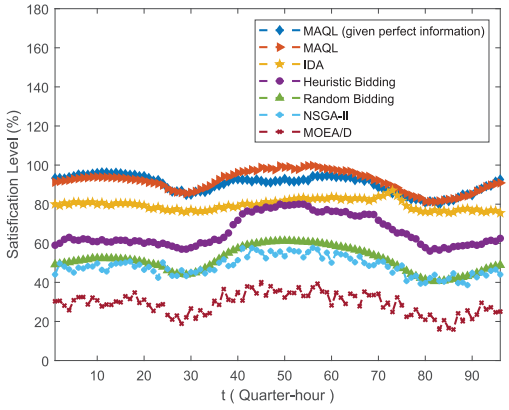
Average numerical results from Fig. 8 are summarized as follows. For the renewable generators, the proposed approach outperformed IDA by 28.74%, heuristic bidding by 32.55%, random bidding by 20.64%, NSGA-II-based approach by



(a)



(b)



(c)

Fig. 10. Performance comparison at the renewable generators and EUs given the microgrid controller optimization model of the aggregator. The best but ideal performance was attained by the MAQL given perfect information on renewable generation. In (a) and (b), the proposed MAQL-based bidding scheme achieved highest profits and rewards for renewable generators and EUs; in (c), the MAQL attained approximately a 100% satisfaction level for EUs.

52.89%, and MOEA/D-based approach by 78.42%. For the EUs, the proposed approach outperformed IDA by 29.29%, heuristic bidding by 47.26%, random bidding by 57.54%, NSGA-II-based approach by 64.64%, and MOEA/D-based approach by 87.97%. On average, the proposed bidding strategies were better than IDA by approximately 29%, heuristic bidding by 39.9%, random bidding by 38.1%, NSGA-II-based approach by 62%, and MOEA/D-based approach by 83.1%.

In an elastic market, small changes in prices yield large changes in supply and demand. Fig. 9 illustrates bidding price variations at EUs in which the excessive supply quantity is the difference between the total renewable generation and total demand. The excessive supply quantity is positive if the renewable supply is higher than the demand and is negative if the supply is lower than the demand. For the proposed approach, higher/lower bidding prices occurred when the excessive supply quantity is negative/positive, and a change of the excessive supply quantity is sensitive to a change of prices, presenting the elastic property. For the random bidding, bidding prices were irrelevant to the excessive supply quantity. For multiobjective optimization approaches, price variations did not elastically reflect the difference between the supply and demand.

Finally, to illustrate that the proposed approach is not restricted to a particular model of the aggregator, we replaced the profit maximization model in (27) with a microgrid controller optimization model [26], [27]. Fig. 10 presents the performance comparison. The proposed MAQL implemented at renewable generators and EUs was able to learn superior bidding strategies than the comparable approaches. This indicates that our learning-based bidding approach should be applicable regardless of the aggregator's business model.

V. Conclusion

This article proposed a learning-based bidding strategy for renewable energy trading in double-sided auctions. A real-time renewable market consisting of an aggregator, renewable generators, and EUs was examined. Renewable generators and EUs were considered as agents aiming to maximize their individual utility functions. Owing to the lack of information about the business model of the aggregator and other agents' bidding profiles, learning problems were formulated and the MAQL-based solution method was developed accordingly. Information about the power quantity bought or sold and the associated prices was then used to update Q-tables. The agents employed their Q-tables that represent the cumulative reward for price bidding in two ways: to explore possible better pricing strategies, the agents determine the prices uniformly at random from time to time; to exploit the available knowledge, the agents greedily select prices that maximize the Q-values most of the time. The main findings from our numerical analysis are as follows.

1) The learning-based bidding strategies were not sensitive to the learning rate once the rate was below a reasonable threshold.
2) The bidding strategies fully exploited state information about aggregate renewable supply and demand.

3) The learning-based framework can readily generalize to include function approximation that addresses a large state dimension, if necessary.
4) The bidding strategies outperformed iterative double auction by approximately 29%, heuristic bidding by 39.9%, random bidding by 38.1%, NSGA-II-based approach by 62%, and MOEA/D-based approach by 83.1% on average.

The major benefit of using our model-free approach is that it is general and can be applied to various bidding scenarios given minimal information. By contrast, existing model-based approaches, for example, must use models about other agents' bidding profiles, assume an oligopolistic market, or assume that agents are merely price takers. As such, several future areas for research based on the proposed methodology are possible, including the consideration of risk-constrained bidding in microgrids. However, learning-based bidding incurs time cost during the learning process. Although an off-policy method was used, implying that any heuristic but effective bidding approaches can be used while the Q-tables are updated in order to reduce the time cost, such a combination of heuristic and learning-based bidding strategies needs further investigation to justify its effectiveness.

REFERENCES

[1] M. Pilz and L. Al-Fagih, "Recent advances in local energy trading in the smart grid based on game-theoretic approaches," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1363–1371, Mar. 2019.

[2] M. M. Esfahani, A. Hariri, and O. A. Mohammed, "A multiagent-based game-theoretic and optimization approach for market operation of multimicrogrid systems," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 280–292, Jan. 2019.

[3] S. Wang, A. F. Taha, J. Wang, K. Kvaternik, and A. Hahn, "Energy crowdsourcing and peer-to-peer energy trading in blockchain-enabled smart grids," *IEEE Trans. Syst.*, *Man, Cybern., Syst.*, vol. 49, no. 8, pp. 1612–1623, Aug. 2019.

[4] K. Zhang *et al.*, "Incentive-driven energy trading in the smart grid," *IEEE Access*, vol. 4, pp. 1243–1257, 2016.

[5] W. Tushar, T. K. Saha, C. Yuen, D. Smith, and H. V. Poor, "Peer-to-peer trading in electricity networks: An overview," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3185–3200, Jul. 2020.

[6] S. Burger, J. P. Chaves-Ávila, C. Batlle, and I. J. Pérez-Arriaga, "A review of the value of aggregators in electricity systems," *Renew. Sustain. Energy Rev.*, vol. 77, pp. 395–405, Sep. 2017.

[7] L. Gkatzikis, I. Koutsopoulos, and T. Salonidis, "The role of aggregators in smart grid demand response markets," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1247–1257, Jun. 2013.

[8] P. Gope and B. Sikdar, "An efficient privacy-friendly hop-by-hop data aggregation scheme for smart grids," *IEEE Syst. J.*, vol. 14, no. 1, pp. 343–352, Mar. 2020.

[9] F. Ahmad, M. S. Alam, and M. Shahidehpour, "Profit maximization of microgrid aggregator under power market environment," *IEEE Syst. J.*, vol. 13, no. 3, pp. 3388–3399, Sep. 2019.

[10] F. Khnlenz, P. H. J. Nardelli, and H. Alves, "Demand control management in microgrids: The impact of different policies and communication network topologies," *IEEE Syst. J.*, vol. 12, no. 4, pp. 3577–3584, Dec. 2018.

[11] Y. Okajima, K. Hirata, T. Murao, T. Hatanaka, V. Gupta, and K. Uchida, "Strategic behavior and market power of aggregators in energy demand networks," in *Proc. IEEE Conf. Decis. Control*, Melbourne, Australia, Dec. 2017, pp. 694–701.

[12] Z. Li, L. Chen, and G. Nan, "Small-scale source trading: A contract theory approach," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1491–1500, Apr. 2018.

[13] D. An, Q. Yang, W. Yu, X. Yang, X. Fu, and W. Zhao, "SODA: Strategy-proof online double auction scheme for multimicrogrids bidding," *IEEE Trans. Syst.*, *Man, Cybern., Syst.*, vol. 48, no. 7, pp. 1177–1190, Jul. 2018.

[14] D. Li, Q. Yang, W. Yu, D. An, Y. Zhang, and W. Zhao, "Towards differential privacy-based online double auction for smart grid," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 971–986, Aug. 2019.

[15] A. Mehdizadeh and N. Taghizadegan, "Robust optimisation approach for bidding strategy of renewable generation-based microgrid under demand side management," *IET Renewable Power Gener.*, vol. 11, no. 11, pp. 1446–1455, 2017.

[16] J. Wang *et al.*, "Optimal bidding strategy for microgrids in joint energy and ancillary service markets considering flexible ramping products," *Appl. Energy*, vol. 205, pp. 294–303, Nov. 2017.

[17] M. N. Faqiry and S. Das, "Double-sided energy auction in microgrid: Equilibrium under price anticipation," *IEEE Access*, vol. 4, pp. 3794–3805, 2016.

[18] N. Li, L. Chen, and M. A. Dahleh, "Demand response using linear supply function bidding," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1827–1838, Mar. 2015.

[19] H. Wu, M. Shahidehpour, A. Alabdulwahab, and A. Abusorrah, "A game theoretic approach to risk-based optimal bidding strategies for electric vehicle aggregators in electricity markets with variable wind energy resources," *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 374–385, Jan. 2016.

[20] W. Wei, F. Liu, and S. Mei, "Energy pricing and dispatch for smart grid retailers under demand response and market price uncertainty," *IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1364–1374, Dec. 2015.

[21] E. Nekouei, T. Alpcan, and D. Chattopadhyay, "Game-theoretic frameworks for demand response in electricity markets," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 748–758, Nov. 2015.

[22] N. Rezaei, A. Ahmadi, A. Khazali, and J. Aghaei, "Multiobjective risk-constrained optimal bidding strategy of smart microgrids: An IGDT-based normal boundary intersection approach," *IEEE Trans. Ind. Informat.*, vol. 15, no. 3, pp. 1532–1543, Mar. 2019.

[23] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, Massachusetts, London, England: MIT Press, 2018.

[24] D. Li and S. K. Jayaweera, "Machine-learning aided optimal customer decisions for an interactive smart grid," *IEEE Syst. J.*, vol. 9, no. 4, pp. 1529–1540, Dec. 2015.

[25] J. Hou *et al.*, "An energy imbalance settlement mechanism considering decision-making strategy of retailers under renewable portfolio standard," *IEEE Access*, vol. 7, pp. 118146–118161, 2019.

[26] N. Azizan Ruhi, K. Dvijotham, N. Chen, and A. Wierman, "Opportunities for price manipulation by aggregators in electricity markets," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5687–5698, Nov. 2018.

[27] M. N. Faqiry and S. Das, "Double auction with hidden user information: Application to energy transaction in microgrid," *IEEE Trans. Syst.*, *Man, Cybern., Syst.*, vol. 49, no. 11, pp. 2326–2339, Nov. 2019.

[28] Universal Smart Energy Framework. [Online]. Available: https://www.usef.energy/

[29] A. Gosavi, "Reinforcement learning: A tutorial survey and recent advances," *INFORMS J. Comput.*, vol. 21, no. 2, pp. 178–192, 2009.

[30] R. Bellman, *Dynamic Programming*. New York, NY, USA: Dover, 2003.

[31] D. E. Kirk, *Optimal Control Theory: An Introduction*. New York, NY, USA: Dover, 2012.

[32] S. Chen and R. S. Cheng, "Operating reserves provision from residential users through load aggregators in smart grid: A game theoretic approach," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1588–1598, Mar. 2019.

[33] D. Li, W.-Y. Chiu, H. Sun, and H. V. Poor, "Multiobjective optimization for demand side management program in smart grid," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1482–1490, Apr. 2018.

[34] N. K. Yadav, M. Kumar, D. Sharma, A. Bala, and G. Bhargava, "Development of bidding strategies using genetic algorithm in deregulated electricity market," in *Proc. Int. Conf. Control, Comput.*, Commun. Materials, Allahabad, India, Oct. 2016, pp. 1–5.

[35] German electricity market. [Online]. Available: https://www.smard.de/en

[36] L. Gigoni *et al.*, "Day-ahead hourly forecasting of power generation from photovoltaic plants," *IEEE Trans. Sustain. Energy*, vol. 9, no. 2, pp. 831–842, Apr. 2018.

[37] Y. Tang, J. Ling, T. Ma, N. Chen, X. Liu, and B. Gao, "A game theoretical approach based bidding strategy optimization for power producers in power markets with renewable electricity," *Energies*, vol. 10, no. 5, May 2017, Art. no. 627.

[38] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Aug. 2002.

[39] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Trans. Evol. Comput.*, vol. 11, no. 6, pp. 712–731, Nov. 2007.

Wei-Yu Chiu (Member, IEEE) received the Ph.D. degree in communications engineering from National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2010.

He is currently an Associate Professor of electrical engineering with NTHU. His research interests include multiobjective optimization and reinforcement learning, and their applications to control systems, robotics, and smart energy systems.

Dr. Chiu was the recipient of the Youth Automatic Control Engineering Award bestowed by Chinese Automatic Control Society in 2016, the Outstanding Young Scholar Academic Award bestowed by Taiwan Association of Systems Science and Engineering in 2017, the Erasmus+Programme Fellowship funded by European Union (staff mobility for teaching) in 2018, and Outstanding Youth Electrical Engineer Award bestowed by Chinese Institute of Electrical Engineering in 2020. From 2015 to 2018, he had been serving as an Organizer/Chair for the International Workshop on Integrating Communications, Control, and Computing Technologies for Smart Grid (ICT4SG). He is a Subject Editor for IET Smart Grid.

Chan-Wei Hu received the B.S. and M.S. degrees in electrical engineering from National Tsing Hua University, Hsinchu, Taiwan. He is currently working toward the Ph.D. degree in computer science at Texas A&M University, Texas, USA.

His research interests include machine learning and computer architectures.

Kun-Yen Chiu received the B.S. degree in electrical engineering from Northeastern University, Boston, MA, USA, in 2018. He is currently working toward the M.S. degree in electrical engineering with National Tsing Hua University, Hsinchu, Taiwan.

His research interests include reinforcement learning algorithms and smart grids.